REMARKS/ARGUMENTS

This amendment responds to the office action dated November 13, 2006.

The applicant has broadened each of independent claims 1, 29, and 56 to recite the limitations of a first and second "semantic characterization of a said play" (claim 1) or "event" (claims 29, 56) to substitute for the earlier recitation of a "type of semantic event," which could have been erroneously been interpreted to preclude more than one semantic characterization of the same play/event. For example, a play/event in a basketball game can both be characterized as a fast-break score and a slam-dunk score. The prior language may have been misinterpreted to preclude the independent claims from reading on an embodiment where that fast break, slam dunk score was indicated, for example, by a red (fast break) blinking (slam dunk) bar or other graphic. In the same vein, the applicant has amended dependent claim 2 to recite the additional limitation of "wherein said first and second semantic characterizations of a said play temporally overlap in said summarization."

The Examiner rejected claims 1-7, 9-34, 36-56, and 58-86 under 35 U.S.C. § 103(a) as being unpatentable over the combination of Christel, "Adjustable Filmstrips and Skims as Abstractions for a Digital Video Library" IEEE Advances in Digital Libraries Conference, May 1999 (hereinafter Christel), in view of Vasconcelos et al., "Bayesian Modeling of Video Editing and Structure: Semantic Features for Video Summarization and Browsing"(hereinafter Vasconcelos) and in further view of Ahmad et al., U.S. Patent No. 6,880,171 (hereinafter Ahmad). The applicant respectfully argues that the Examiner's rejection is improper, as independent claims 1, 29, and 56 each recite limitations that patentably distinguish over the cited prior art.

With respect to independent claim 1, that claim recites the limitations of (1) "summarizing a video . . . based upon an event characterized by a semantic event that includes a play"; (2) "displaying a graphical user interface . . . said interface sequentially indicating the relative location of each of said plurality of segments within said summarization relative to at least one other of said segments"; and (3) "displaying [a] relative location for a first type of semantic characterization of said play in said video using a first visual indication and displaying said relative location for a second type of semantic characterization of a said play in said video

using a second visual indication different from said first visual indication." Only the second of these recited limitations are disclosed in the cited prior art. The Examiner contends that the first limitation is obvious in view of Christel, and that the third limitation is obvious in view of the combination of Christel, Vasconcelos, and Ahmad. Both contentions are incorrect.

Christel discloses a system for presenting video skims in which a user may enter a specific query to which certain frames of a video are "matched." The video skim is constructed by (1) using a query from a user to identify matching key frames in a selected video, and (2) based upon those matching frames, constructing a summarization that builds video segments around each of the matching frames. The first of these steps is accomplished by matching words in the query to words in descriptive text of the video, constructed from either a speech recognition module or annotations to the video, such as close captioning, or both. *See, e.g.* Christel at p. 3 col. 2. Because the query-matching module of Christel relies so heavily upon a textual description of the video, Christel discloses that the system is limited to an Informedia video collection that includes news and documentaries, i.e. genres of video for which the textual descriptions not only actually describe the content of what is visually presented in the video, but is timed to coincide with the segments they describe. *See, e.g.* Christel at p. 1 col. 1 section 1, par. 1 (describing the system being applied to news and documentary videos); *See also Id.* at p. 3 col. 2, section 3 par. 3 (stating that the retrieval engine relies upon matching words in a query to descriptive text "timed tightly" to the video segment). Therefore, the Examiner's assertion at page 4 of the present office action that "Christel includes all types of video when discussing capturing important events from a video" and that "when the video is of a sporting event, the essential content would be plays" lacks support in the actual disclosure of Christel. To the contrary, not only does Christel specifically limit the disclosure to types of video *other than* sporting events, but it would seem to one of ordinary skill in the art that the video skim construction module of Christel would be inappropriate for sports video, where it cannot be reliably assumed that video segments showing a particular semantic type of play, e.g. a "slam dunk", "fast break", etc. would be usually accompanied by synchronous ones of either close-captioning or audio saying "slam dunk", "fast break", etc. Thus, the conclusion of the Examiner, that the limitation of "summarizing a video . . . based upon an event characterized by a semantic

event that includes a play", is but an obvious variant of the system of Christel lack support in the prior art.

Moreover, the Examiner's conclusion that it would be obvious, in light of Vasconcelos and Ahmad, to modify Christel to arrive at the remaining limitations of claim 1 also lacks support in the prior art. At the outset, and even setting aside whether the events summarized by Christel could include "plays", to the extent that Christel does disclose "summarizing a video, said summarization comprising a plurality of segments of said video based upon an event", the "event" that forms the basis for selecting the plurality of segments to include in the summary is unknown to all but the contemporaneous user, hence is both unpredictable, and already tailored to the user's specific interests. In fact, Christel touts these features:

> Just as we modified filmstrips so that match locations were taken into account, so video skims were adjusted from early work to emphasize the audio and video surrounding match locations. Rather than being pre-computed, these new style video skims are generated dynamically so that context can be used to assemble better skims, e.g. following a query the skim will be assembled to emphasize the locations in the video where match locations are found.

See Christel at p. 4 col. 2 lines 38-47.

This primary reference notes, however, that merely constructing a summary of a video by extracting the segments in a video that best match a user's contemporaneous interests is insufficient to hold the user's interest in the skim; mere extraction of these segments tended to produce an aesthetically displeasing, choppy and unsynchronized video. *See* Christel at p. 4 col. 2 lines 16-31. To improve the fluidity of the summary presentation, Christel proposed a method of constructing a skim by expanding segments around match locations, where the length of each segment was determined by a combination of (1) user input as to the compression ratio for the summary as a whole; and (2) "goodness values" calculated for automatically-generated segment cutpoints. *See Id.* at p. 4 col. 2 line 45 to p. 5 col. 1 line 4. Specifically, given user input of a query and a desired compression ratio,

> [t]he skim is initialized to consist of sequences containing any of the given match locations, merging sequences which occur very close together. The sequences in the skim are then expanded: the sequence endpoint with the worst goodness rating is extended out to the next cutpoint, thus embedding that bad cutpoint into the skim. This process repeats until the target skim size is reached.

See Id. at p. 5 col. 1 lines 14-21.

Given the inherent trade-off between the user-selected compression ratio and the fluidity of the skim presentation, Christel shows a user interface that roughly communicates to a user the marginal benefit of decreasing the selected compression ratio (increasing the length of the summary), and allows a user to adjust the compression ratio accordingly. In FIG. 5, for example, Christel shows a user interface that, in addition to playing the desired skim, shows two bars. The first bar indicates the relative location of matching frames in the video being skimmed, while the second bar indicates the relative location of the segments automatically constructed around those matching frames. From these two bars, a user can estimate the marginal improvement in presentation by decreasing the selected compression ratio. For example, in FIG. 5 a majority of the match locations to the exemplary query are found near the beginning of the video, and many of the segments are interrupted by only a short interval. Thus, it would be reasonable to assume that marginally decreasing the compression ratio would achieve a proportionally greater benefit in presentation fluidity. This is confirmed by FIG. 6, where, by decreasing the amount of compression from 20% (5:1 compression) to 40% (5:2 compression) the number of breaks between segments was reduced from 12 to 5. Moreover, FIG. 6 shows a slider allowing the viewer to incrementally adjust the compression ratio using feedback from the segment and match point location bars.

Therefore, although Christel discloses the claimed step of "displaying a graphical user interface on a second portion of said display, said interface sequentially indicating the relative location of each of said plurality of segments within said summarization relative to at least one other of said segments as each of said plurality of segments is displayed" as recited in independent claim 1, Christel does so solely for the purpose of providing statistical feedback to the user as to the marginal benefits received in exchange for the cost of further decreasing the compression ratio, increasing the length of the summary. The graphical user interface of Christel is not intended to distinguish the relative temporal locations of different types of semantic content, nor would it be used for such a purpose because the summary of Christel is already constructed in response to a specific user inquiry as to the type of semantic content the user wishes to see.

In fact, the Examiner concedes that Christel's user interface does not teach the claimed limitation of "displaying said relative location for a first type of *semantic event* in said video

using a first visual indication and displaying said relative location for a second type of *semantic event* in said video using a second visual indication different from said first visual indication." Instead, the Examiner asserts that Vasconcelos and Ahmad provide the motivation to modify Christel to provide this missing limitation. Neither reference, however, either teaches this limitation or provides a motive for modifying Christel to provide this limitation.

Vasconcelos teaches a method of automatically identifying a high-level semantic domain of a film, i.e. whether the film is of a particular genre, e.g. action, romance, etc. Drawing on the observation that different genres of movies have different image characteristics, e.g. that dramas tend to be heavy on dialogue and close-ups of actors' faces, while action films tend to employ fast cuts with fewer close-ups of actors, Vasconcelos describes that a video may be characterized by four timelines, shown in FIG. 2, that each show the temporal locations of respective ones of four characteristic *statistical* types of video *shots*, i.e. close-up shots, fast-cut shots, crowd shots, and natural settings. *See* Vasconcelos, FIG. 2; *see also Id.* at p. 153, section 1, par. 2 (stating that the method identifies the structure of shots in a video from which semantic attributes can be inferred). Vasconcelos discloses that by comparing these timelines for a video, a viewer can *infer* the semantic type of film being shown. Stated differently, Vasconcelos discloses the use of timelines, not for distinguishing among a plurality of types of semantic events in a video, but instead for distinguishing among a plurality of types of visual composition of shots. *See* Vasconcelos at p. 154 sec. 2 (describing a Bayesian inference process where detected *structural features* of movies are mapped to a graph, from which inferences can be made as to semantic content). The genre of film can then be inferred from viewing the timelines together, e.g. if the timelines show more close-up shots than fast-cut scenes (both characteristics of the shot composition, as opposed to a type of semantic event in the video itself), then a viewer can infer a likelihood that the video is of the drama genre than the action genre.

With this in mind, the Examiner's assertion at p. 4 lines 11-13 of the present office action, that "Vasconcelos discloses identifying specific semantic events and displaying identified data through visual indications in a timeline" is facially incorrect. Not only do the timelines of Vasconcelos fail to display specific semantic events in a video, but Vasconcelos seems to disclaim that the system disclosed therein is capable of doing so. *See* Vasconcelos at p. 153 section 1 paragraph 4 (describing the practical outcome of the theoretical approach outlined in

the paper as being "generic", i.e. only characterizing the *domain* of movies); *See also Id.* at p. 156, section 5.3 ("The characterization is not fine enough to [automatically] distinguish between The River Wild and Ghost and the Darkness"). Therefore neither Vasconcelos nor the primary reference, Christel, discloses the limitation of "displaying said relative location for a first *semantic characterization of a said play* in said video using a first visual indication and displaying said relative location for a second *semantic characterization of a said play* in said video using a second visual indication different from said first visual indication."

Nor does the tertiary reference, Ahmad, disclose this limitation. Ahmad discloses a browser for audiovisual content where a user can view summary information related to available content. In a specific embodiment, noted by the Examiner, Ahmad discloses a window showing, as an example, "news programs" available for viewing where any currently viewed news program is shaded in one color while news programs that have already been viewed are shaded in another color. *See* Ahmad at col. 16 lines 54-65. Presumably, were the window showing "action movies" or "documentaries" the window could be similarly marked to shade, for example, any currently viewed documentary one color and previously viewed documentaries another color. Thus, the different colored shadings, as taught by Ahmad, are not indicative of any semantic content *in* the video; rather, the differing visual indications are merely indicative of the *statistical property* of whether that viewer is either currently watching the program (shading in one color), has previously watched the program (shading in another color), or neither (no shading). The applicant further notes that the post-facto marking of content as being either watched or not watched cannot indicate anything meaningful about the events in a video created long before the user had the opportunity to watch the program.

The term "semantic event" relates to the *meaning* of an event, and more specifically, the claim limitation of a "semantic characterization of a play" (or event) in a video relates to a meaning of a particular play or event portrayed. For example, if the video is of a basketball game, a type of semantic characterization of a play (event) in the video might include slam dunks, fast breaks, fouls, and injuries. If the video is an action movie, types of semantic characterizations of events in the video might include car chases, explosions, and gunfights. Even a cursory reading of Ahmad shows that it fails to disclose the limitation of "displaying said relative location for a first semantic characterization of a play (or event) in said video using a

first visual indication and displaying said relative location for second semantic characterization of a play (or event) in said video using a second visual indication different from said first visual indication."

In view of the respective disclosures of Christel, Vasconcelos, and Ahmad, each previously described, the Examiner's rejection of independent claim 1 as being obvious over the combination of these references is deficient on any one of several grounds. First, as previously explained, none of the cited references disclose using visual indicia in a graphical interface to indicate respective types of *semantic content* depicted in a video being summarized. Instead, each reference uses visual indicia to show statistical or structural properties of either the video (e.g. Vasconcelos' timelines showing types of structure of shots; Ahmad's colors indicating the statistical feature of whether a video has been watched) or a summary of a video (e.g. Christel's video scroll bars showing match locations and segment locations used in a video summary, relative to the summarized video).

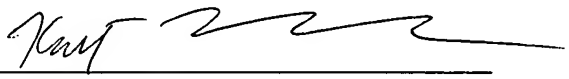
Second, one of ordinary skill in the art would not make the combination suggested. If, for example, the Examiner is arguing for a substitution of Vasconcelos' timelines (as modified to include Ahmad's different colors for different shot-types) for Christel's scroll bars, then such a substitution would frustrate the very purpose of Christel's user interface, which is to provide the user feedback as to the marginal benefit of making the summary a little longer. On the other hand, if the Examiner is suggesting that Christel's user interface be modified to include, in addition to the scroll bars, the timelines of Vasconcelos, the Examiner fails to provide a motive for doing so; a user of Christel's system already knows the genre of the video being summarized, and is being presented with a summary specifically constructed in response to a query as to the type of content desired to be seen. A user of Christel's summary has no need for the timelines of Vasconcelos because there is no need to infer, using a Bayesian model or otherwise, what content is being presented when the content presented already matches a specific query.

Finally, none of the cited references disclose or suggest the limitation of "summarizing a video . . . based upon an event characterized by a semantic event that includes a play." Therefore, for each of these reasons, independent claim 1, as well as its dependent claims 2-7 and 9-28, each patentably distinguishes over the cited prior art.

Each of independent claims 29 and 56 includes the limitation of "displaying said temporal location for a first semantic characterization of an event . . . using a first visual indication and displaying said temporal location for a second type of semantic characterization of an event . . . using a second visual indication different from said first visual indication." Therefore, each of these claims, as well as their respective dependent claims 30-34, 36-55, and 58-86, patentably distinguish over the cited prior art for the same reasons as does independent claim 1.

In view of the foregoing remarks, the applicant respectfully requests reconsideration and allowance of claims 1-7, 9-34, 36-56, and 58-86.

Respectfully Submitted



Kurt Rohlfs
Reg. No. 54,405
Tel: (503) 227-5631